CLAIMS

What is claimed is:

1 1. A method of enterprise web mining comprising the steps of:

2 collecting data from a plurality of data sources;

3 integrating the collected data;

4 generating a plurality of data mining models using the collected data; and

5 generating a prediction or recommendation in response to a received

6 request for a recommendation or prediction.

1 2. The method of claim 1, wherein the collecting step comprises the steps of:

2 acquiring data from the plurality of data sources;

3 selecting data that is relevant to a desired output from among the acquired

4 data;

5 pre-processing the selected data; and

6 building a plurality of database tables from the pre-processed selected

7 data.

1 3. The method of claim 2, wherein the plurality of data sources comprises:

2 proprietary account or user-based data;

3 complementary external data;

4 web server data; and

5   web transaction data.

1   4.   The method of claim 3, wherein the web server data comprises:

2   at least one of: web traffic data obtained by Transmission Control

3   Protocol/Internet Protocol packet sniffing, web traffic data obtained from an

4   application program interface of the web server, and a log file of the web server.

1   5.   The method of claim 2, wherein the acquired data comprises a plurality of

2   different types of data and integration step comprises the step of:

3   forming an integrated database comprising collected data in a coherent

4   format.

1   6.   The method of claim 5, wherein the model generating step comprises the

2   steps of:

3   selecting an algorithm to be used to generate a model;

4   generating at least one model using the selected algorithm and data

5   included in the integrated database; and

6   . deploying the at least one model.

1   7.   The method of claim 6, wherein the step of deploying the at least one

2   model comprises the step of:

-75-

3   generating program code implementing the model.

1   8.   The method of claim 7, wherein the step of generating an online

2   prediction or recommendation comprises the steps of:

3   receiving a request for a prediction or recommendation;

4   scoring a model using data included in the integrated database;

5   generating a predication or recommendation based on the generated score;

6   and

7   transmitting the predication or recommendation.

1   9.   The method of claim 8, wherein the step of pre-processing the selected

2   data comprises the step of:

3   performing, on the selected data, at least one of: data cleaning, visitor

4   identification, session reconstruction, classification of web pages into

5   navigation and content pages, path completion, and converting file names to

6   page titles.

1   10.   The method of claim 8, wherein the step of pre-processing the selected

2   data comprises the step of:

3   collecting pre-defined items of data passed by a web server.

1 11.   A computer program product for performing an enterprise web mining

2 process in an electronic data processing system, comprising:

3   a computer readable medium;

4   computer program instructions, recorded on the computer readable

5 medium, executable by a processor, for performing the steps of:

6   collecting data from a plurality of data sources;

7   integrating the collected data;

8   generating a plurality of data mining models using the collected data; and

9   generating a prediction or recommendation in response to a received

10 request for a recommendation or prediction.

1 12.   The computer program product of claim 11, wherein the collecting step

2 comprises the steps of:

3   acquiring data from the plurality of data sources;

4   selecting data that is relevant to a desired output from among the acquired

5 data;

6   pre-processing the selected data; and

7   building a plurality of database tables from the pre-processed selected

8 data.

1   13.   The computer program product of claim 12, wherein the plurality of data

2   sources comprises:

3   proprietary account or user-based data;

4   complementary external data;

5   web server data; and

6   web transaction data.

1   14.   The computer program product of claim 13, wherein the web server data

2   comprises:

3   at least one of: web traffic data obtained by Transmission Control

4   Protocol/Internet Protocol packet sniffing, web traffic data obtained from an

5   application program interface of the web server, and a log file of the web server.

1   15.   The computer program product of claim 12, wherein the acquired data

2   comprises a plurality of different types of data and integration step comprises the

3   step of:

4   forming an integrated database comprising collected data in a coherent

5   format.

1   16.   The computer program product of claim 15, wherein the model generating

2   step comprises the steps of:

3   selecting an algorithm to be used to generate a model;

4   generating at least one model using the selected algorithm and data

5   included in the integrated database; and

6   deploying the at least one model.

1   17.   The computer program product of claim 16, wherein the step of deploying

2   the at least one model comprises the step of:

3   generating program code implementing the model.

1   18.   The computer program product of claim 17, wherein the step of

2   generating an online prediction or recommendation comprises the steps of:

3   receiving a request for a prediction or recommendation;

4   scoring a model using data included in the integrated database;

5   generating a predication or recommendation based on the generated score;

6   and

7   transmitting the predication or recommendation.

1   19.   The computer program product of claim 18, wherein the step of pre-

2   processing the selected data comprises the step of:

3   performing, on the selected data, at least one of: data cleaning, visitor

4   identification, session reconstruction, classification of web pages into

5   navigation and content pages, path completion, and converting file names to

6   page titles.

1   20.   The computer program product of claim 18, wherein the step of pre-

2   processing the selected data comprises the step of:

3   collecting pre-defined items of data passed by a web server.

1   21.   A system for performing an enterprise web mining process, comprising:

2   a processor operable to execute computer program instructions; and

3   a memory operable to store computer program instructions executable

4   by the processor, for performing the steps of:

5   collecting data from a plurality of data sources;

6   integrating the collected data;

7   generating a plurality of data mining models using the collected data; and

8   generating a prediction or recommendation in response to a received

9   request for a recommendation or prediction.

1   22.   The system of claim 21, wherein the collecting step comprises the steps

2   of:

3   acquiring data from the plurality of data sources;

4 selecting data that is relevant to a desired output from among the acquired

5 data;

6   pre-processing the selected data; and

7   building a plurality of database tables from the pre-processed selected

8 data.

1 23. The system of claim 22, wherein the plurality of data sources comprises:

2   proprietary account or user-based data;

3   complementary external data;

4   web server data; and

5   web transaction data.

1 24. The system of claim 23, wherein the web server data comprises:

2   at least one of: web traffic data obtained by Transmission Control

3 Protocol/Internet Protocol packet sniffing, web traffic data obtained from an

4 application program interface of the web server, and a log file of the web server.

1 25. The system of claim 22, wherein the acquired data comprises a plurality of

2 different types of data and integration step comprises the step of:

3   forming an integrated database comprising collected data in a coherent

4 format.

-81-

1   26.   The system of claim 25, wherein the model generating step comprises the

2   steps of:

3   selecting an algorithm to be used to generate a model;

4   generating at least one model using   the   selected   algorithm   and   data

5   included in the integrated database; and

6   deploying the at least one model.

1   27.   The system of claim 26, wherein the step of deploying the at least one

2   model comprises the step of:

3   generating program code implementing the model.

1   28.   The   system   of   claim   27,   wherein   the   step   of   generating   an   online

2   prediction or recommendation comprises the steps of:

3   receiving a request for a prediction or recommendation;

4   scoring a model using data included in the integrated database;

5   generating a predication or recommendation based on the generated score;

6   and

7   transmitting the predication or recommendation.

1   29.   The system of claim 28, wherein the step of pre-processing the selected

2   data comprises the step of:

3   performing, on the selected data, at least one of: data cleaning, visitor

4   identification, session reconstruction, classification of web pages into

5   navigation and content pages, path completion, and converting file names to

6   page titles.

1   30.   The system of claim 28, wherein the step of pre-processing the selected

2   data comprises the step of:

3   collecting pre-defined items of data passed by a web server.

1   31.   An enterprise web mining system comprising:

2   a database coupled to a plurality of data sources, the database operable to

3   store data collected from the data sources;

4   a data mining engine coupled to the web server and the database, the data

5   mining engine operable to generate a plurality of data mining models using the

6   collected data;

7   a server coupled to a network, the server operable to:

8   receive a request for a prediction or recommendation over the network,

9   generate a prediction or recommendation using the data mining models,

10 and

11 transmit the generated prediction or recommendation.

1 32. The system of claim 31, wherein the database comprises:

2 a plurality of database tables built from the collected data.

1 33. The system of claim 32, wherein the plurality of data sources comprises:

2 proprietary account or user-based data;

3 complementary external data;

4 web server data; and

5 web transaction data.

1 34. The system of claim 33, wherein the web server data comprises:

2 at least one of: web traffic data obtained by Transmission Control

3 Protocol/Internet Protocol packet sniffing, web traffic data obtained from an

4 application program interface of the web server, and a log file of the web server.

1 35. The system of claim 32, wherein the plurality of database tables forms an

2 integrated database comprising collected data in a coherent format.

1 36. The system of claim 35, wherein the data mining engine is further

2 operable to:

3 select an algorithm to be used to generate a model;

4 generate at least one model using the selected algorithm and data included

5 in the integrated database; and

6 deploy the at least one model.

1 37. The system of claim 36, wherein the deployed model comprises program

2 code implementing the model.

1 38. The system of claim 37, wherein the server is operable to generate a

2 prediction or recommendation by scoring a model using data included in the

3 integrated database and generating a predication or recommendation based on the

4 generated score.

1 39. The system of claim 31, further comprising a data pre-processing engine

2 pre-processing the selected data.

1 40. The system of claim 39, wherein the database comprises:

2 a plurality of database tables built from the pre-processed selected data.

1 41. The system of claim 40, wherein the plurality of data sources comprises:

2 proprietary account or user-based data;

3   complementary external data;

4   web server data; and

5   web transaction data.

1   42.   The system of claim 41, wherein the web server data comprises:

2   at least one of: web traffic data obtained by Transmission Control

3   Protocol/Internet Protocol packet sniffing, web traffic data obtained from an

4   application program interface of the web server, and a log file of the web server.

1   43.   The system of claim 40, wherein the plurality of database tables forms an

2   integrated database comprising collected data in a coherent format.

1   44.   The system of claim 43, wherein the data mining engine is further

2   operable to:

3   select an algorithm to be used to generate a model;

4   generate at least one model using the selected algorithm and data included

5   in the integrated database; and

6   deploy the at least one model.

1   45.   The system of claim 44, wherein the deployed model comprises program

2   code implementing the model.

1   46.   The system of claim 45, wherein the server is operable to generate a

2   prediction or recommendation by scoring a model using data included in the

3   integrated database and generating a predication or recommendation based on the

4   generated score.

1   47.   The method of claim 46, wherein the data pre-processing engine pre-

2   processes the selected data by performing, on the selected data, at least one of:

3   data cleaning, visitor identification, session reconstruction, classification of

4   web pages into navigation and content pages, path completion, and converting

5   file names to page titles.

1   48.   The method of claim 47, wherein the data pre-processing engine pre-

2   processes the selected data by collecting pre-defined items of data passed by a

3   web server.